

PEMBINAAN DAN PENGESAHAN INSTRUMEN BAGI MENGUKUR TAHAP LITERASI PENTAKSIRAN GURU SEKOLAH MENENGAH DI MALAYSIA

Rohaya Talib
Mohd Najib Abd Ghafar
Fakulti Pendidikan
Universiti Teknologi Malaysia

Abstrak

Dalam usaha membangunkan sistem pendidikan bertaraf dunia, pengetahuan dan kemahiran pentaksiran seharusnya dipertingkatkan bagi membolehkan guru membantu proses pembelajaran secara berkesan. Sehubungan itu, tujuan kajian adalah membina dan mengesah instrumen - Ujian Literasi Pentaksiran (ULP) bagi mengukur kompetensi pengetahuan pentaksiran guru dari lima dimensi: (i) konsep pentaksiran, (ii) kaedah pengukuran, (iii) pengujian, (iv) penskoran dan penggredan dan (v) statistik dan pelaporan. ULP mengandungi 60 item aneka pilihan polikotomos berasaskan Kurikulum Pengukuran dan Penilaian Program Latihan Keguruan. Data ULP dianalisis menggunakan program Quest - gabungan pendekatan Classical Test Theory (CTT) dan Model Rasch dan penilaian item berpandukan indeks diskriminsi, threshold, Infit MeanSquare dan analisis distraktor. Indeks kebolehppercayaan (KR_{20}) menunjukkan peningkatan dari .71(ULP-R1) kepada .85 (ULP-R2). Hasil kajian mendapati ULP mempunyai ciri-ciri psikometriks yang baik dan diterima dalam kajian pembinaan instrumen. Justeru, kajian telah dapat mengenalpasti tahap literasi pentaksiran guru-guru sekolah menengah di Malaysia dan norm berdasarkan faktor jantina, jawatan, pengalaman mengajar dan bidang mata pelajaran.

1.0 Pengenalan

Pelan Induk Pembangunan Pendidikan telah dilancarkan sekitar Februari 2007. Fokus PIPP bagi Sistem Pentaksiran Pendidikan Kebangsaan (SPPK) adalah menyediakan pentaksiran alternatif, mengkaji semula sistem peperiksaan dan memantapkan kualiti sistem pentaksiran secara keseluruhan. Lembaga Peperiksaan Malaysia sedang menyusun semula kerangka dan format pentaksiran peringkat nasional dan merancang memantapkan aktiviti Pentaksiran Berasaskan Sekolah (PBS). Selaras dengan perubahan fokus dan format pentaksiran, kompetensi para guru perlu dipertingkatkan agar dapat menjalankan tugas secara profesional dan berwibawa. Pajares (1992) dan Title (1994) mengenalpasti salah satu faktor mempengaruhi keberkesanan pelaksanaan pentaksiran di sekolah adalah pengetahuan dan kemahiran guru dalam bidang pentaksiran pendidikan. Black dan William (1998) mengesahkan literasi pentaksiran merupakan kriteria penting ke arah menjayakan matlamat pendidikan berkualiti. Sehubungan itu, beberapa kajian berkaitan amalan pentaksiran (Ismail Alias, 1993; Salbiah Ishak, 1998; Muhamad Sahari Nordin, 2001; Mohd Azhar Mat Ali, 2006) telah mencadangkan agar guru diberikan latihan berterusan dalam bidang pentaksiran. Justeru, pembinaan instrumen bertujuan mengukur tahap literasi pentaksiran guru adalah signifikan. Hasil pengukuran dapat menjelaskan kekuatan dan kelemahan pengetahuan pentaksiran guru dan dijadikan panduan dalam merancang program pembangunan profesionalisme.

2.0 Literasi Pentaksiran

Konsep literasi pentaksiran merujuk kepada kompetensi berasaskan pengetahuan dalam bidang pentaksiran (Schaefer, 1993; Stiggins, 1999; Mertler, 2005) merangkumi pengetahuan deklaratif (mengenai fakta, konsep, prinsip – *knowing what*) dan pengetahuan prosedural (aplikasi pengetahuan kepada amalan- *knowing how*) (Chen Jiamu, 2001).

Assessment literates know the difference between sound and unsound assessment. They know what they are assessing, why they are doing it, how best to assess the knowledge or skill of interest, also aware of the potential negative consequences of poor and inaccurate assessment.

(Stiggins, 1995; 240)

Bagi kajian ini, pengkaji mengemukakan definisi konseptual literasi pentaksiran sebagai kompetensi pengetahuan guru dalam bidang pentaksiran. Berasaskan pengetahuan, guru dapat mengamalkan prinsip, tatacara, strategi dan kaedah pentaksiran yang sesuai bagi membantu proses pembelajaran seterusnya membuat keputusan yang tepat dan adil mengenai pelajar, kurikulum dan program pendidikan secara keseluruhan. Jadual 1.1 menjelaskan konstruk pengetahuan pentaksiran mengikut dimensi.

Jadual 1 Konstruk pengetahuan pentaksiran

Bil	Dimensi	Objektif
1	Konsep Pentaksiran	<ul style="list-style-type: none"> • Menghuraikan tujuan dan peranan pentaksiran dalam pengajaran dan pembelajaran • Menjelaskan kepentingan pentaksiran sekolah • Menerangkan istilah yang berkaitan dengan pentaksiran • Mengenalpasti ciri-ciri utama ujian • Membandingbezakan jenis-jenis ujian • Membandingbezakan jenis-jenis penilaian
2	Kaedah Pengukuran	<ul style="list-style-type: none"> • Mengenal pasti kaedah pengukuran bersesuaian dengan objektif pengajaran [kaedah penulisan, pemerhatian, lisan].
3	Pengujian	<ul style="list-style-type: none"> • Memahami pengertian, tujuan, ciri-ciri utama dan jenis-jenis ujian • Menyatakan tatacara pembinaan Jadual Spesifikasi Ujian [isi kandungan, format dan jenis item] • Menyatakan kepentingan Jadual Spesifikasi Ujian dalam pembinaan ujian • Menghuraikan prosedur membina item objektif dan subjektif • Menentukan kualiti item • Menerangkan langkah pentadbiran ujian [peringkat persediaan, pelaksanaan dan penyelarasan]
4	Penskoran & Penggredan	<ul style="list-style-type: none"> • Mengaplikasikan penskoran secara analitik dan holistik • Menerangkan tatacara dan kepentingan skema penskoran • Memahami keperluan analisis item dan bank item dalam

		pengendalian penilaian <ul style="list-style-type: none"> • Menerangkan prosedur pengredan [asas pengredan dan tafsiran]
5	Statistik & Pelaporan	<ul style="list-style-type: none"> • Memahami konsep asas statistik (skor mentah, kekerapan, jeda kelas, graf : poligon, histogram, ogif) • Menguasai kemahiran memindahkan skor mentah kepada graf spt poligon, histogram dan ogif • Menggunakan rumus-rumus perangkaan untuk membuat tafsiran skor • Menganalisis dan menginterpretasi data untuk menyediakan laporan. • Menerangkan tujuan, kaedah dan kegunaan laporan

4.0 Kepentingan Literasi Pentaksiran

Guru yang kompeten pentaksiran dapat merancang strategi pengajaran dengan berkesan selain mampu meningkatkan motivasi pelajar (Stiggins, 1997; Brookhart, 1999). Kurang pengetahuan pentaksiran menyebabkan guru gagal memantau kemajuan pembelajaran (Gallagher, 1998), tidak dapat menghasilkan keputusan yang adil (Gronlund & Linn, 1990; Hills, 1991; Stiggins, 2001), perancangan pengajaran yang lemah (Stiggins, 1991; Brookhart, 1999) dan tidak dapat mencungkil potensi sebenar pelajar (Stiggins, 1995; Airasian, 2000; Stiggins, 2001). Guru didapati melaporkan kepada ibu bapa dan pelajar secara tidak tepat (Plake, 1993) mengakibatkan ketidakpercayaan (McKenna, 1977) dan keraguan (Herman & Golan, 1993). Seterusnya, kualiti pentaksiran dipersoalkan oleh pelbagai pihak disebabkan akauntabiliti pentaksiran gagal dilaksanakan sepenuhnya (Stiggins, 2001; Mertler, 2005).

4.0 Kajian Tentang Literasi Pentaksiran

Sorotan penulisan menunjukkan tidak terdapat instrumen yang boleh digunakan bagi mengukur tahap literasi pentaksiran guru di Malaysia. Di Barat, literasi pentaksiran mula dikaji sekitar tahun 1990an apabila polisi desentralisasi pentaksiran mengambil tempat dalam sistem pendidikan (Webb, 2002). Mandat mempertingkatkan peranan pentaksiran sekolah bagi mencemerlangkan prestasi ujian piawai mengakibatkan guru-guru dipertanggungjawabkan sepenuhnya atas kejayaan atau kegagalan pelajar (Mertler, 2005). Justeru, isu literasi pentaksiran guru mula dibincang dan dikaji. Terdapat laporan kajian mengesahkan bahawa guru tidak dilatih secukupnya dalam pentaksiran (Gullickson, 1984), kurikulum latihan pentaksiran tidak memenuhi keperluan pentaksiran bilik darjah (Schaefer & Lissitz, 1987) dan kompetensi pentaksiran guru berada pada tahap lemah (Brookhart, 2001; Campbell *et al.*, 2002). Sehubungan itu, Stiggins (1999) mengesa isu literasi pentaksiran dikaji kerana ia menghalang proses perkembangan diri pelajar.

Plake *et al.*, (1993) telah memulakan kajian literasi pentaksiran dengan membina ujian *Classroom Assessment Literacy Inventory (CALI)*. CALI terdiri dari 35 item berteraskan tujuh prinsip *Standards for Teachers Competence in Educational Assessment of Student*. Ia ditadbir ke atas 555 orang guru dari 98 daerah dan 45 negeri.

Indeks kebolehppercayaan $KR_{20} = 0.54$ dan skor purata 23 daripada 35 (66%). Hasil kajian mendapati pencapaian guru berpengalaman lebih baik berbanding guru kurang pengalaman. Min skor tertinggi ujian adalah bagi standard 3 (pentadbiran, penskoran dan pentafsiran) manakala min skor paling rendah adalah Standard 6 (menyampaikan hasil pentaksiran kepada pelajar, ibu bapa dan pihak yang berkaitan).

Campbell *et al.* (2002) meneruskan kajian menggunakan ujian sama terhadap 220 orang siswazah yang telah melengkapkan kursus pengukuran dan penilaian. Darjah kebolehppercayaan KR_{20} 0.74 dan min skor 21 daripada 35 (60%). Kesan langsung pembelajaran telah meningkatkan indeks kebolehppercayaan instrumen. Namun, skor purata ujian 60% didapati lebih rendah berbanding kajian Plake *et al.* (1993). Min skor tertinggi pada standard 1 (memilih kaedah pentaksiran bersesuaian dengan pengajaran) berbanding standard 3 pada kajian sebelumnya manakala min skor paling rendah sama bagi kedua-dua kajian iaitu Standard 6 (menyampaikan hasil pentaksiran kepada pelajar, ibu bapa dan pihak yang berkaitan). Perbandingan dapatan menunjukkan pengalaman dan latihan mempengaruhi tahap literasi pentaksiran.

Fokus kajian Mertler (2003) adalah membandingkan tahap literasi guru dalam perkhidmatan dengan guru pra-perkhidmatan. Dapatan kajian menunjukkan nilai psikometriks instrumen hampir sama dengan kajian Plake *et al.* (1993) dan Campbell *et al.* (2002). Mertler (2005) pula telah menambahbaik instrumen dari aspek kejelasan bahasa, kebolehbacaan dan ketepatan pilihan jawapan dan dikenali sebagai *Assessment Literacy Inventory* (ALI). ALI dirintis sebanyak dua kali ke atas guru pra-perkhidmatan seramai 152 orang dari dua buah institusi pengajian tinggi. Hasil kajian rintis seperti berikut : Indeks kebolehppercayaan $KR_{20} = 0.74$, Indeks diskriminasi = 0.31, min skor = 22.98 dan sisihan piawai = 4.05. Keputusan kajian menunjukkan ALI adalah satu alat ukur yang mempunyai ciri-ciri kesahan dan kebolehppercayaan yang tinggi. Nilai kebolehppercayaan yang konsisten iaitu 0.74 adalah nilai yang digalakkan oleh Kehoe (1995), Chase (1999) dan Nitko (2001). Dapatan kajian mereka mendapati bahawa matlamat kursus pentaksiran tidak selaras dengan keperluan pentaksiran di bilik darjah dan latihan keguruan tidak memberi pengetahuan dan kemahiran pentaksiran secukupnya kepada guru. Justeru, Mertler (2005) mencadangkan pengetahuan dan kemahiran pentaksiran dipertingkatkan secara berterusan dalam konteks program pembangunan profesionalisme guru.

5.0 Pernyataan Masalah

Pentaksiran adalah satu mekanisme atau proses pengumpulan maklumat bertujuan meningkatkan kualiti pengajaran dan pembelajaran (Simmons & Resnick, 1993; Arter, 2003). Keberkesanan pelaksanaannya bergantung kepada tahap literasi pentaksiran guru (Stiggins, 1995; Mertler, 2003). Sehingga kini, tahap literasi pentaksiran guru tidak dapat diukur kerana ketiadaan alat yang dapat digunakan bagi tujuan tersebut. Dengan itu, tujuan utama kajian ini adalah membina instrumen Ujian Literasi Pentaksiran yang sah dan dipercayai bagi tujuan mengukur tahap literasi pentaksiran guru-guru sekolah menengah di Malaysia.

6.0 Objektif Kajian

Kajian ini bertujuan:

1. membina instrumen Ujian Literasi Pentaksiran (ULP) dengan ciri-ciri
 - (a) kebolehpercayaan
 - (b) kesahan
2. mengukur tahap literasi pentaksiran guru sekolah menengah di Malaysia
 - (a) menentukan norm berdasarkan faktor jantina, jawatan, pengalaman mengajar dan bidang mata pelajaran.
 - (b) melakar profil berdasarkan faktor jantina, jawatan, pengalaman mengajar dan bidang mata pelajaran.

7.0 Metodologi Kajian

7.1 Reka bentuk kajian

Kajian ini bersifat kuantitatif dengan pengumpulan data dirancang dalam tiga fasa menggunakan pendekatan tinjauan deskriptif : Fasa I Kajian Keperluan, Fasa II : Pembinaan Instrumen dan Fasa III : Pengesahan Instrumen (Rujuk Lampiran 1).

7.2 Populasi dan sampel kajian

Populasi kajian berjumlah 136 598 orang guru-guru sekolah menengah di seluruh Malaysia. Namun, populasi diandaikan *homogeneous*; responden terdiri dari guru-guru yang berkhidmat di sekolah-sekolah menengah kerajaan dan menerima input kurikulum yang standard semasa program latihan keguruan. Tambahan lagi, aktiviti dan tugas pentaksiran adalah seragam bagi semua sekolah di Malaysia. Atas sifat *homogeneous* tersebut, pemilihan sampel dilakukan secara rawak berkelompok di kalangan guru-guru di Negeri Johor (17 186 orang guru). Sejumlah 501 sampel dari 167 buah sekolah dari 10 daerah dipilih dengan tahap keyakinan 95% berdasarkan anggaran 4% ralat sampel (Fowler, 1988).

7.3 Instrumen kajian

Tiga instrumen digunakan dalam kajian. Dua instrumen (SS1 dan SS2) semasa fasa kajian keperluan dan satu instrumen (versi ULP, ULP-R1 dan ULP-R2) bagi fasa pembinaan dan fasa pengesahan. Item ULP dirancang pembinaannya dengan menyediakan Jadual Spesifikasi Ujian berdasarkan lima dimensi pengetahuan iaitu (i) konsep pentaksiran-KOP, (ii) kaedah pengukuran-KAP, (iii) pengujian-PUJI, (iv) penskoran dan penggredan-PSG dan (v) statistik dan pelaporan-STP. Langkah pembinaan ULP berpandukan cadangan McIntire dan Miller (2007). Item disemak melalui dua prosedur iaitu (i) kualitatif (*test-taker review* dan *experts review*) dan (ii) kuantitatif (analisis item) menggunakan perisian Quest. Set ULP mengandungi tiga bahagian iaitu Bahagian A : Maklumat Demografi, Bahagian B : ULP dan Bahagian C : Kertas Jawapan.

7.4 Prosedur pengumpulan data

Data kajian keperluan dikumpul dari 69 orang setiausaha peperiksaan (SS1) dan 28 orang pentadbir sekolah (SS2). Sebaik menerima kelulusan menjalankan kajian, ujian rintis pertama (ULP-R1) ditadbir ke atas 71 orang guru dari 15 buah sekolah dari tiga daerah. Seterusnya, ULP-R2 ditadbir ke atas sampel terdiri dari 465 orang guru dari 155 buah sekolah dari 10 daerah untuk tujuan pengesahan.

7.5 Penganalisan data kajian

Data kajian dianalisis menggunakan perisian komputer. Data kajian keperluan (SS1 dan SS2), data demografi, norm dan profil dianalisis menggunakan MS Excel dalam bentuk frekuensi, peratus dan graf. Perisian Quest digunakan bagi menganalisis item ULP-R1 dan ULP-R2. Output Quest mengandungi indikator seperti Indeks Diskriminasi, *Threshold*, Distraktor dan *Infit MeanSquare*. Perisian SPSS (Anova) digunakan bagi ujian perbezaan min bagi faktor jantina, jawatan, pengalaman mengajar dan bidang mata pelajaran.

7.6 Analisis kualitatif

Item ULP telah disemak oleh enam orang guru (pengalaman mengajar melebihi 10 tahun) dan panel pakar. Guru-guru dikehendaki menjawab kesemua item dan memberikan maklumbalas dari aspek iaitu kejelasan maksud, bahasa dan masa. Selepas disemak panel pakar, lima item (18,23,26,31 dan 46) dibuang kerana tidak mendapat persetujuan 100%. Pengguguran lima item dari set ULP telah merubah jumlah item bagi setiap dimensi seperti berikut :

KOP : **7** [1,2,3,19,26,47,50] kepada **6** [1,2,3,40,42,45]
 KAP : **7** [9,10,15,34,37,48,56] kepada **8** [9,10,15,18,30,33,43,51]
 PUJI : **26** [4,5,6,8,13,21,22,24,28,29,30,31,35,36,39,40,41,44,46,51,52,53,54,55,58,60]
 kepada **21** [4,5,6,8,13,20,21,22,25,26,27,32,36,37,41,46,47,48,49,50,53]
 PSG : **5** [14,17,18,20,57] kepada **7** [14,16,17,19,31,35,52]
 STP : **15** [[7,11,12,16,23,25,27,32,33,38,42,43,45,49,59] kepada **13**
 [7,11,12,23,24,28,29,34,38,39,44,54,55]

7.7 Kajian rintis 1

Sampel kajian terdiri dari 71 orang guru dari 15 buah sekolah dari tiga daerah. Data dikumpul dan dianalisis menggunakan program SPSS dan Quest. Jadual 1.2 menunjukkan statistik deskriptif ULP-R1.

Jadual 2 Statistik deskriptif ULP-R1

Pengiraan Ukuran Memusat		Pengukuran Ukuran Serakan	
Min	22.93	Julat	6-35
Median	23	Varians	39.95
Mod	27	Sisishan Piawai	6.28

Indeks Kebolehpercayaan $KR_{20} = 0.71$

Item dipilih berdasar indikator *Infit MeanSquare*, *Threshold*, Indeks Diskriminasi dan analisis distraktor dengan syarat-syarat berikut:

- *Infit MeanSquare* (IMS) dalam julat 0.77 - 1.30
- *Threshold* (Indeks kesukaran) dalam julat - 2.0 - + 2.0
- Indeks Diskriminasi (ID) dalam julat 0.20 – 0.8
- Setiap distraktor dipilih minimum 3% dan peratus memilih distraktor tidak melebihi peratus jawapan

Semua item ULP didapati berada dalam julat IMS antara 0.85 hingga 1.27 dan julat *Threshold* antara -1.26 hingga 1.87. Walau bagaimanapun, 19 item menunjukkan ID berada antara 0 dan 0.19 manakala 9 item menunjukkan nilai ID negatif. Justeru, 27 item diterima manakala 19 item perlu diubahsuai (4 item - stimulus dan 15 item - distraktor) dan 9 item diubahsuai keseluruhan (indeks diskriminasi negatif).

7.8 Kajian Rintis 2

Sampel kajian terdiri dari 465 orang guru sekolah menengah dari 155 buah sekolah dari 10 daerah. Maklumat demografi ULP-R2 seperti berikut :

- Jantina – Lelaki 33.3%, Perempuan 66.7%
- Jawatan – Guru Mata Pelajaran 33.3%, Ketua Panitia 33.3%, Setiausaha Peperiksaan 33.3%
- Pengalaman Mengajar – 1-5 Tahun 23.4%, 6-10 Tahun 20.2%, Lebih 10 Tahun 56.3%
- Bidang Mata Pelajaran – Bahasa 24.7%, Sains & Matematik 44.3%, Teknik & Vokasional 11.4%, Kemanusiaan 19.6%.

Statistik deskriptif ULP-R2 ditunjukkan dalam Jadual 1.3.

Jadual 3 Statistik deskriptif ULP-R2

Pengiraan Ukuran Memusat		Pengukuran Ukuran Serakan	
Min	28.71	Julat	8-48
Median	30	Varians	70.83
Mod	32	Sisihan Piawai	8.41
Indeks Kebolehppercayaan $KR_{20} = 0.84$			

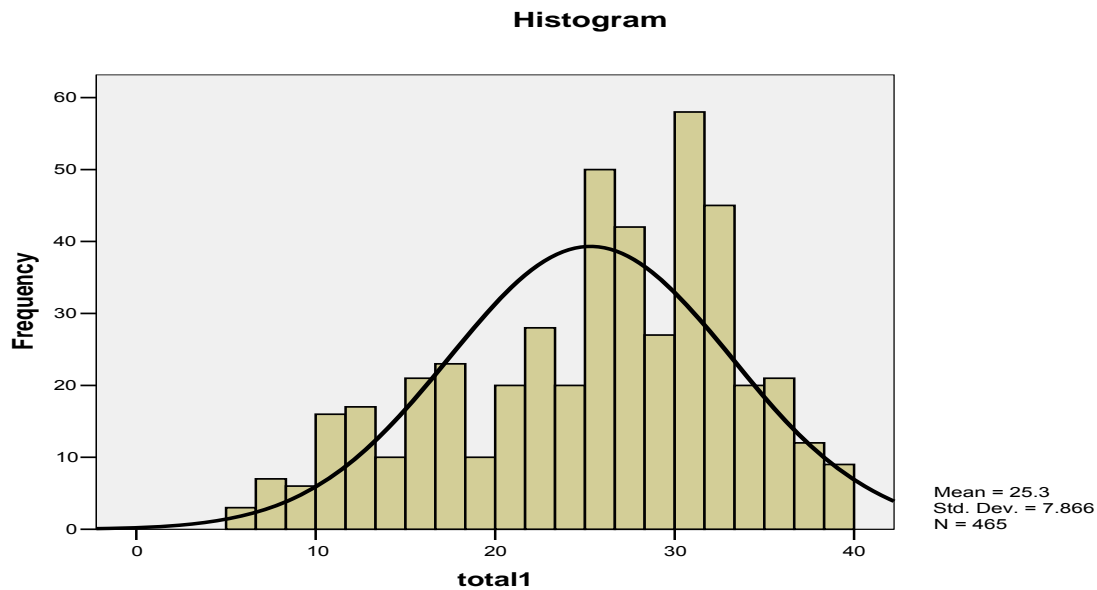
Jadual 3 menunjukkan kualiti item ULP-R2 bertambah baik di mana indeks kebolehppercayaan meningkat dari 0.71 kepada 0.84. Matriks item menunjukkan semua item ULP didapati berada dalam julat IMS antara 0.81 hingga 1.16 dan julat *Threshold* antara -1.60 hingga 1.86. Daripada 55 item, satu item (8) menunjukkan ID negatif, tujuh item (1, 15,23,29,32,44,54) menunjukkan ID berada dalam julat antara 0 dan 0.19 manakala dua item (34,37) menunjukkan peratus pemilihan distraktor melebihi peratus jawapan. Pengkaji membuat keputusan untuk menghapuskan 10 item tersebut kerana tidak memenuhi syarat pemilihan yang telah ditetapkan. Dengan itu, set ULP-R2 mengandungi 45 item dianalisis untuk kali kedua. Jadual 1.4 menunjukkan statistik deskriptif ULP-R2(2).

Jadual 4 Statistik deskriptif ULP-R2(2)

Pengiraan Ukuran Memusat		Pengukuran Ukuran Serakan	
Min	25.30	Julat	5-35
Median	27	Varians	61.87
Mod	31	Sisihan Piawai	7.86

Indeks Kebolehppercayaan $KR_{20} = 0.85$

Maklumat dari Jadual 4 dan Rajah 1 menunjukkan min skor ULP 25.30 dengan sisihan piawai 7.86. Skor tertinggi sampel adalah 35 manakala terendah 5. Memandangkan nilai median taburan hampir kepada min, maka taburan secara keseluruhannya dikatakan menghampiri normal. Lengkok taburan skor normal membuktikan bahawa sampel-sampel yang dipilih mewakili ciri-ciri populasi kajian (Black, 2002).

**Rajah 1** Taburan skor ULP

8.0 Dapatan Kajian

8.1 Analisis kesahan dan kebolehppercayaan

Instrumen ULP dihasilkan dengan darjah kebolehppercayaan KR_{20} yang tinggi iaitu 0.85 dan didapati mengukur satu konstruk (*unidimensionality*) dibuktikan melalui indeks *Infit MeanSquare* keseluruhan berada pada nilai 1.00 (Sisihan Piawai .09) dengan anggaran kebolehppercayaan item berdasarkan *Threshold* pada nilai .96. Peratus pendiskriminasi item agak seimbang iaitu kumpulan diskriminasi tinggi 30%, kumpulan diskriminasi sederhana tinggi 36% dan kumpulan sederhana rendah 34%.

Jadual 5 Korelasi dimensi dengan skor total

Dimensi	Koefisien Korelasi (skor keseluruhan)	Tafsiran Alias Baba (1992)
KOP	0.63	Tinggi
KAP	0.70	Tinggi
PUJI	0.89	Sangat Tinggi
PSG	0.71	Tinggi
STP	0.72	Tinggi

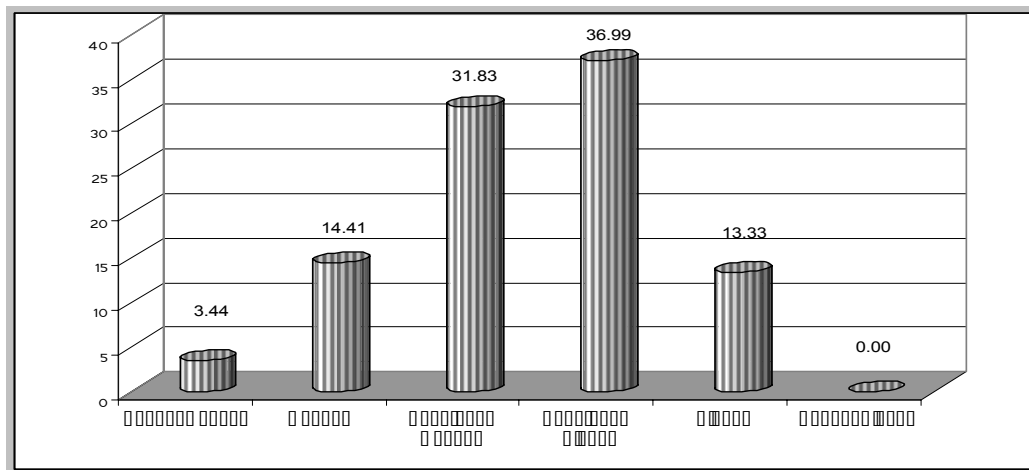
Selain *Infit MeanSquare*, analisis korelasi *Pearson Product Moment* digunakan bagi menentukan kesahan konstruk. Jadual 1.5 menunjukkan koefisien korelasi positif antara dimensi dengan skor total dalam julat 0.63 – 0.89 (tinggi hingga sangat tinggi) pada aras signifikan .01. Korelasi tinggi antara dimensi dengan skor total memberi petanda baik bagi kesahan konstruk (Kline, 2002). Sebaliknya, koefisien korelasi antara dimensi dengan dimensi lain berada dalam julat sederhana dan rendah (0.57 hingga 0.34). Hubungan positif dengan nilai korelasi kecil menunjukkan setiap dimensi mempunyai perkaitan dalam mengukur konstruk pengetahuan pentaksiran.

8.2 Tahap literasi pentaksiran guru sekolah menengah di Malaysia

Jadual 6 Indikator tahap literasi pentaksiran

Tahap	Indikator	Skor
Sgt Tinggi	Min + 3 SP	70 < Skor T
Tinggi	Min + 2 SP	60 < Skor T ≤ 70
Sed Tinggi	Min +1 SP	50 < Skor T ≤ 60
Sed Rendah	Min – 1 SP	40 < Skor T ≤ 50
Rendah	Min – 2 SP	30 < Skor T ≤ 40
Sgt Rendah	Min – 3 SP	Skor T ≤ 30

Skor mentah ULP ditukar kepada unit pengukuran standard bagi tujuan perbandingan (McIntire & Miller, 2007). Kajian memilih skor T (min: 50, SP:10) bagi tujuan tersebut. Setelah melalui proses transformasi, tahap literasi pentaksiran guru ditentukan berdasarkan indikator dalam Jadual 1.6. Taburan skor T menunjukkan 0 % (0/465) guru berada pada tahap literasi sangat tinggi, 13.33% (62/465) pada tahap tinggi, 36.99% (172/465) pada tahap sederhana tinggi, 31.83% (148/465) pada tahap sederhana rendah, 14.41% (67/465) pada tahap rendah dan 3.44% (16/465) pada tahap sangat rendah. Sila rujuk Rajah 2.



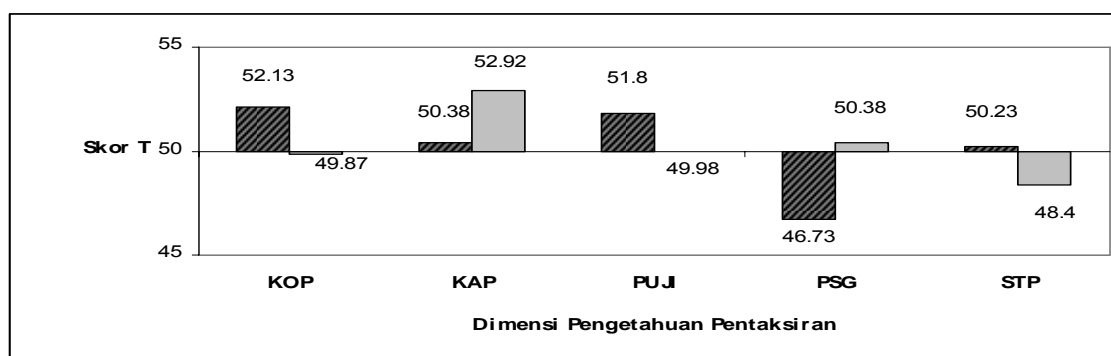
Rajah 2 Tahap literasi pentaksiran guru sekolah menengah di Malaysia

8.3 Norm dan profil

Norm merujuk kepada set skor yang digunakan sebagai asas perbandingan sewaktu proses pengukuran dijalankan (Airasian, 2000). Data deskriptif (min dan sisihan piawai) yang dihasilkan dari sampel kajian membentuk norm secara keseluruhan dan norm bagi sub-konstruk (Lowenthal, 2001). Ia berfungsi sebagai rujukan perbandingan dan tafsiran (Nunnally, 1978). Profil adalah persembahan grafik min skor ujian dibandingkan dengan norm kajian (Brown, 1983). Profil berbentuk graf bar dihasilkan dalam kajian ini. Bahagian 6.3 menerangkan norm dan profil bagi faktor jantina, jawatan, pengalaman mengajar dan bidang mata pelajaran.

8.4 Norm dan profil jantina (2 kumpulan: Lelaki-L, Perempuan-P)

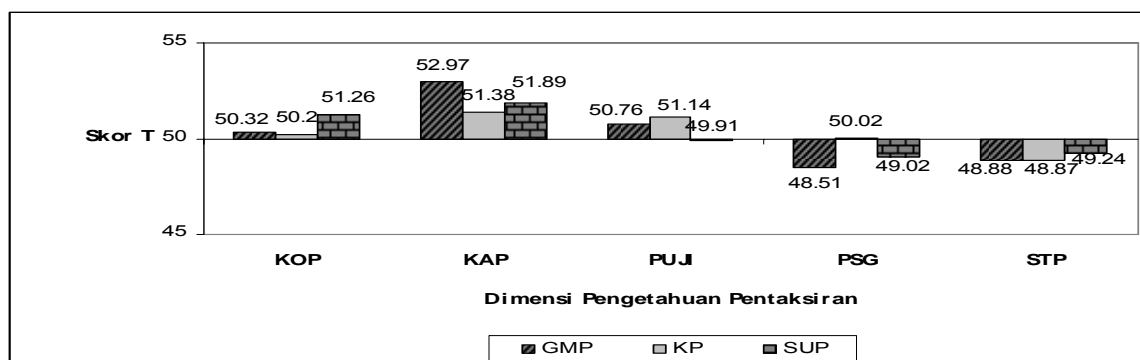
Hasil kajian mendapati norm guru perempuan lebih tinggi (Min 50.15; SP 9.81) berbanding norm guru lelaki (Min 49.69; SP 10.41). Berdasarkan profil jantina mengikut dimensi pengetahuan (Rajah 1.3), norm guru lelaki paling tinggi berada pada dimensi KOP (52.13) dan paling rendah adalah PSG (46.73) manakala norm guru perempuan paling tinggi pada dimensi KAP (52.92) dan paling rendah STP (48.40).



Rajah 3 Profil dua kumpulan jantungina

Norm dan Profil Jawatan (3 kumpulan: Guru Mata Pelajaran-GMP, Ketua Panitia-KP, Setiausaha Peperiksaan-SUP).

Analisis perbezaan min (*Analyze> Compare Means*) menunjukkan norm SUP (Min 50.97; SP 9.65) paling tinggi berbanding norm KP (Min 49.75; SP10.28) diikuti norm GMP (Min 49.29; SP 10.01). Rajah 4 menunjukkan profil tiga kumpulan jawatan.

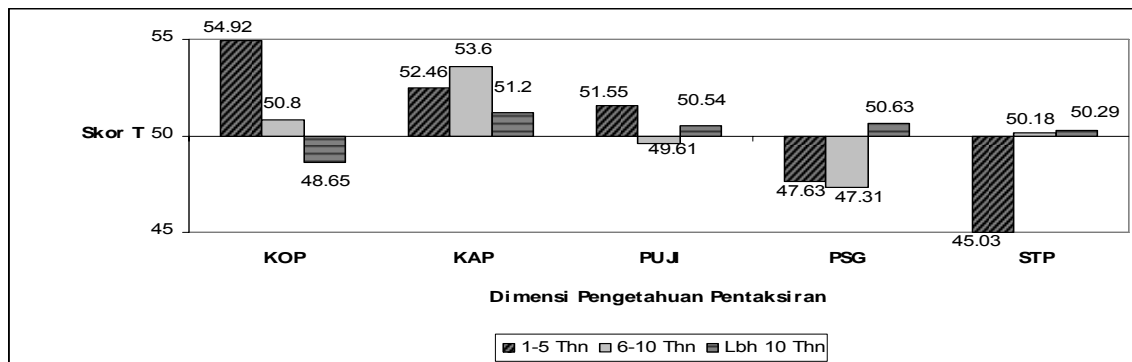


Rajah 4 Profil tiga kumpulan jawatan

Norm paling tinggi bagi GMP adalah dimensi KAP(52.97) manakala paling rendah PSG(48.51). Sama seperti GMP, dimensi paling tinggi bagi KP adalah KAP (51.38) dan paling rendah STP (48.88). Norm SUP paling tinggi juga pada dimensi KAP(51.89) manakala paling rendah PSG (49.02). Ketiga-tiga kumpulan menunjukkan pencapaian di atas norm kajian pada dua dimensi KOP dan KAP manakala dimensi STP dibawah norm kajian.

8.5 Norm dan profil pengalaman mengajar (3 kumpulan : 1-5 Thn (PM1), 6-10 Thn (PM2), Lebih 10 Thn (PM3))

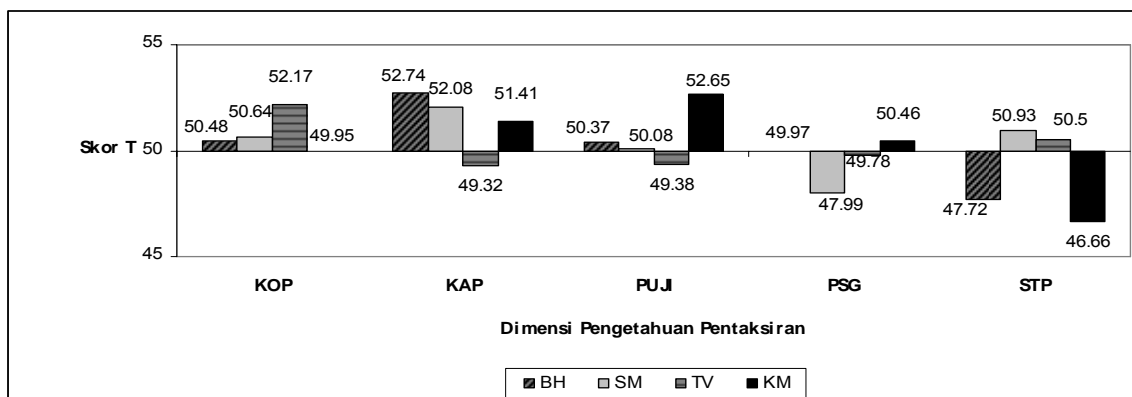
Analisis perbezaan min (*Analyze> Compare Means*) mendapati norm PM2 (Min 50.11; SP 9.74) paling tinggi berbanding PM3 (Min 50.09; SP10.34) dan PM1(Min 49.70; SP9.43). Rajah 1.5 menunjukkan profil tiga kumpulan pengalaman mengajar mengikut dimensi pengetahuan. Norm PM1 paling tinggi pada dimensi KOP(54.92) dan paling rendah dimensi STP (45.03) manakala PM2 menunjukkan norm paling tinggi berada pada dimensi KAP(53.60) dan paling rendah dimensi PSG(47.31). Norm paling tinggi bagi PM3 adalah KAP (51.20) dan terendah KOP(48.65).



Rajah 5 Profil tiga kumpulan pengalaman mengajar

8.6 Norm dan profil bidang mata pelajaran (4 kumpulan: Bahasa-BH, Sains dan Matematik-SM, Teknik dan Vokasional-TV, Kemanusiaan-KM)

Hasil analisis mendapati norm Bidang SM (Min 50.89; SP10.70) paling tinggi diikuti dengan Bidang TV (Min 49.85; SP10.12), Bidang KM (Min 49.66; SP 8.92) dan Bidang BH (Min 48.76; SP 9.38). Seterusnya, Rajah 1.6 menunjukkan profil empat kumpulan pengalaman mengajar berdasarkan dimensi pengetahuan pentaksiran.

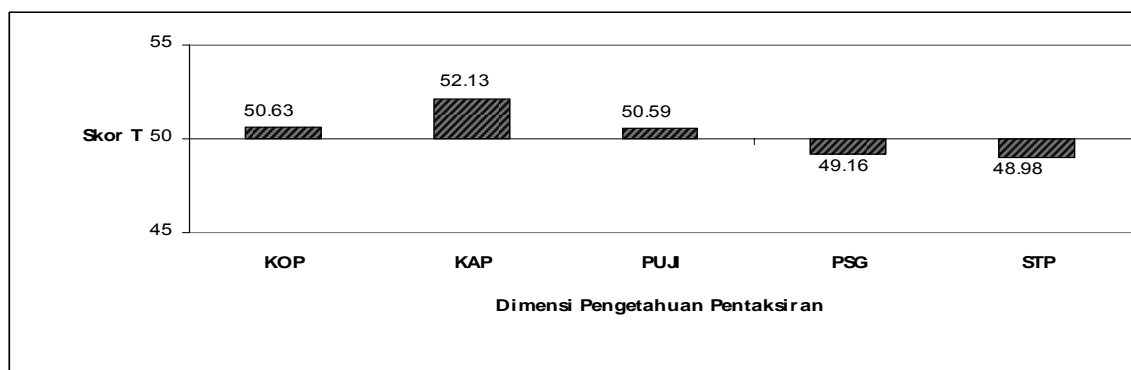


Rajah 6 Profil empat kumpulan bidang mata pelajaran

Bidang BH menunjukkan norm paling tinggi pada dimensi KAP (52.74) dan paling rendah dimensi PDG (49.97). Norm paling tinggi Bidang SM pada dimensi KAP (52.08) manakala dimensi terendah PSG (47.99). Bidang TV menunjukkan norm tertinggi pada dimensi KOP (52.17) dan terendah dimensi KAP (49.32). Norm tertinggi Bidang KM adalah PUJI (52.65) dan terendah dimensi STP(46.66).

8.7 Norm dan profil keseluruhan

Rajah 7 di bawah menunjukkan profil lima dimensi pengetahuan pentaksiran. Norm tertinggi pada dimensi KAP (52.15) diikuti KOP(50.65), PUJI (50.59), PSG (49.16) dan terendah STP(48.98).



Rajah 7 Profil pengetahuan pentaksiran guru-guru sekolah menengah di Malaysia

9.0 Perbincangan

Kajian ini bertujuan membina instrumen - Ujian Literasi Pentaksiran (ULP) bagi mengukur tahap literasi pentaksiran guru sekolah menengah di Malaysia. ULP setelah dirintis menunjukkan darjah kebolehpercayaan dengan nilai .85; satu nilai yang diterima oleh Popham (1990), Kehoe (1995), Nitko (2001) dan Kubiszyn dan Borich (2007). Pendekatan CTT dan Rasch membolehkan pemilihan item dibuat dengan lebih teliti. Bagi kesahan konstruk, analisis Rasch menganggarkan nilai kebolehpercayaan item berdasarkan *Threshold* (Indeks kesukaran) berada pada nilai .96. Nilai *Infit MeanSquare* antara 0.81-1.13 (Min =1.00, SP = 0.09) membuktikan data kajian bersesuaian dengan model dan mengukur satu konstruk yang jelas (*unidimensionality*) (Adams & Khoo, 1993). Analisis tambahan korelasi *Pearson Product Moment* (r) menunjukkan darjah korelasi berada dalam julat tinggi - sangat tinggi bagi setiap dimensi dengan skor total (0.63-0.89; $\alpha = .01$). Sebaliknya, nilai r positif antara dimensi dalam julat sederhana - rendah (0.57 hingga 0.34, $\alpha = .01$) membuktikan item mengukur dimensi berbeza tetapi mempunyai perkaitan positif dengan konstruk (Kline, 2002). Dapatan tersebut membuktikan ULP adalah instrumen yang mempunyai ciri kesahan konstruk dan darjah kebolehpercayaan yang tinggi.

Skor mentah tidak dapat memberikan maklumat berguna bagi menilai pencapaian (Kline, 2002). Justeru, skor T (Min 10; SP 10) digunakan sebagai asas perbandingan dan tafsiran (Airasian, 2000). Analisis skor T menunjukkan tahap literasi guru berada dalam kontinum sederhana (68.82%), tinggi (13.33%) dan rendah (17.85%). Keputusan ini dihasilkan dari jumlah sampel yang dipilih secara rawak dengan tahap keyakinan 95% bagi anggaran 4% ralat sampel (Fowler, 1988). Ini bermakna, dapatan sampel kajian dapat digeneralisasikan kepada ciri populasi (tahap literasi pentaksiran guru sekolah menengah di Malaysia) dengan keyakinan 95% bahawa 4 dari 100 kali min sampel berbeza dari min populasi (Creswell, 2002). Dengan itu, dapatan kajian ini telah dapat mengenalpasti tahap pengetahuan pentaksiran guru sekolah menengah di Malaysia. Sejumlah 86.67% (68.82+17.85) guru berada di tahap sederhana dan rendah. Oleh itu, satu usaha konstruktif perlu dilakukan bagi meningkatkan pengetahuan pentaksiran guru dengan berfokuskan pentaksiran berasaskan sekolah dan pentaksiran alternatif. Mc Munn *et al.*(2003) mencadangkan

agar pembangunan profesionalisme pentaksiran diadakan secara berterusan dan berasaskan model konstruktivis (kolaboratif dan bersepadu).

Analisis norm menunjukkan terdapatnya perbezaan min antara kumpulan bagi setiap faktor. Contohnya faktor jantina menunjukkan perbezaan norm antara guru lelaki (49.69) dan guru perempuan (50.15). Bagi menghasilkan inferens yang tepat, ujian ANOVA berasaskan nisbah F iaitu perbandingan varian antara kumpulan sebagai numerator dengan varian dalam kumpulan sebagai denominator (Othman Mohamed, 2001) dijalankan. Keputusan ujian menunjukkan tidak terdapat perbezaan min yang signifikan antara kumpulan bagi keempat-empat faktor yang dikaji pada aras signifikan .05 (Jantina : $p = .591 > \alpha$, $F(1,463) = .289$; Jawatan : $p = .311 > \alpha$, $F(2,462) = 1.172$, Pengalaman Mengajar : $p = .937 > \alpha$, $F(2,462) = .066$; Bidang Mata Pelajaran : $p = .318 > \alpha$, $F(3,461) = 1.177$). Dengan kata lain, perbezaan min yang terhasil antara kumpulan adalah terlalu kecil jika dibandingkan dengan apa yang dijangkakan berlaku secara kebetulan (Kerlinger & Lee, 2000). Ini bermakna pengetahuan pentaksiran guru tidak berbeza secara signifikan berdasarkan faktor jantina, jawatan, pengalaman mengajar dan bidang mata pelajaran. Dapatan ini didapati bercanggah dengan dapatan Plake *et al.*, (1993) dan Campbell *et al.*, (2002) iaitu pengalaman mengajar dan latihan (khususnya jawatan SUP) mempengaruhi tahap literasi pentaksiran guru.

10.0 Kesimpulan

Ujian Literasi Pentaksiran (ULP) yang dibina didapati mempunyai ciri-ciri psikometriks yang baik setelah diketengahkan bukti-bukti dari sudut kesahan dan kebolehpercayaan. Dengan instrumen tersebut, tahap dan norm literasi pentaksiran telah dapat dikenalpasti berdasarkan data empirikal dan dapatan ini amat berguna untuk memahami kompetensi pengetahuan pentaksiran guru. Dapatan kajian membolehkan program pembangunan profesionalisme dirancang dengan lebih berkesan. Instrumen ULP dapat digunakan oleh organisasi di bawah Kementerian Pelajaran Malaysia khususnya sekolah, Jabatan Pendidikan Daerah dan Jabatan Pelajaran Negeri sebagai ujian kesediaan (*readiness test*) dalam mengenalpasti kekuatan dan kelemahan guru secara khusus mengikut dimensi pengetahuan. Selain itu, dapatan kajian dapat digunakan bagi merancang kurikulum pengukuran dan penilaian dalam program latihan keguruan. Kualiti guru yang dihasilkan adalah bergantung kepada kerelevanan dan keberkesanan pengisian kurikulum semasa latihan keguruan. Kurikulum yang diikuti oleh bakal guru mestilah mampu memberikan pengetahuan dan kemahiran agar mereka bersedia menerima segala tugas dan tanggungjawab apabila memulakan perkhidmatan di sekolah (Robiah Sidin, 1998).

Rujukan

- Adams, R.J. dan Khoo, S. (1993). *Quest : The Interactive Test Analysis System*. Hawthorn, Victoria : ACER.
- Airasian, P. W. (2000). *Assessment in the Classroom : A Concise Approach*. (2nd Ed.). Boston : McGraw Hill.
- Alias Baba (1999). *Statistik Penyelidikan dalam Pendidikan dan Sains Sosial*. Bangi : Penerbit Universiti Kebangsaan Malaysia.
- Arter, J. A. (2003). *Assessment for Learning: Classroom Assessment to Improve Student Achievement and Well-Being*. ERIC : US Department of Education.

- Black, P. dan William, D. (1998). Inside the Black Box : Raising Standards Through Classroom Assessment. *Phi Delta Kappan*. October, 139-148.
- Black, T.R. (2002). *Understanding Social Science Research* (2nd ed.). London : Sage Publications Ltd.
- Brookhart, S. M. (1999). *The Art and Science of Classroom Assessment : The Missing Part of Pedagogy*. Washington DC ERIC Clearinghouse on Higher Education and Office of Educational Research and Improvement.
- Brookhart, S. M. (2001). The Standard and Classroom Assessment Research. *Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education, Dallas, TX*. (ERIC Document Reproduction Service No. ED 451189).
- Brown, F. G. (1983). *Principles of Educational and Psychological Testing*. (3rd ed.). New York : Holt, Rinehart and Winston, Inc.
- Campbell, C., Murphy, J.A. dan Holt, J. K. (2002). Psychometric Analysis of an Assessment Literacy Instrument : Applicability to Preservice Teachers. *Paper presented at the Annual Meeting of the Mid-Western Educational Research Association*. October, 2006. Columbus, OH.
- Chase, C. I. (1999). *Contemporary Assessment for Educators*. New York : Addison Wesley Longman.
- Chen Jiamu (2001). The Great Importance of the Distinction Between Declarative and Procedural Knowledge. *Analise Psicologica*. 4(21), 559-566.
- Creswell, J. W. (2002). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. Upper Saddle River , NJ : Merrill Prentice Hall.
- Fowler, F. J. (1988). *Survey Research Methods*. Newbury Park, CA : Sage.
- Gallagher, J. D. (1998). *Classroom Assessment for Teachers*. Upper Saddle River, New Jersey : Merrill Prentice Hall.
- Gronlund, N. E. dan Linn, R.L. (1990). *Measurement and Evaluation in Teaching* (6th ed.). Upper Saddle River, NJ : Merrill-Prentice Hall.
- Gullickson, A. R. (1984). Teacher Perspectives on Their Instructional Use of Tests. *Journal of Educational Research*, 77, 244-248.
- Herman, J.L., dan Golan, S. (1993). The Effects of Standardized Testing on Teaching and Schools. *Educational Measurement : Issues and Practice*, 12(4), 20-25.
- Hills, J.R. (1991). Apathy Concerning Grading and Testing. *Phi Delta Kappan*, 72, 540-545.
- Ismail Alias (1993). Tanggapan Terhadap Penilaian Formatif, Pembinaan Item Penilaian dan Penggunaanya oleh Guru Sains dan Matematik. Disertasi Sarjana Pendidikan. Universiti Kebangsaan Malaysia.
- Kehoe, J. (1995). Basic Item Analysis for Multiple-Choice Tests. *Practical Assessment, Research and Evaluation*, 4(10).
- Kerlinger, F. N., dan Lee, H.B. (2000). *Foundations of Behavioral Research* (4th ed.). United States : Wadsworth Thomson Learning.
- Kline, T. (2005). *Psychological Testing : A Practical Approach to Design and Evaluation*. Thousand Oaks, California : Sage Publications, Inc.
- Lowenthal, K. M. (2001). *An Introduction to Psychological Tests and Scales*(2nd ed.). University of London : Psychology Press.
- Mcintire, S.A. dan Miller, L.A. (2000). *Foundations of Psychological Testing*. New York : McGraw Hill.
- McKenna, B. (1977). What's Wrong with Standardized Testing? *Today's Education*, 36.

- McMunn, N., McColskey, W. dan Butler, S. (2004). Building Teacher Capacity in Classroom Assessment To Improve Student Learning. *International Journal of Educational Policy, Research & Practice*, 4(4), 25-48.
- Mertler, C. A. (2003). Preservice Versus Inservice Teachers' Assessment Literacy : Does Classroom Experience Make a Difference? *Paper presented at the Annual Meeting of the Mid-Western Educational Research Association*. Oct 15-18, 2003. Columbus, OH.
- Mertler, C. A. (2005). Measuring Teachers' Knowledge and Application of Classroom Assessment Concepts : Development of the Assessment Literacy Inventory. *Paper presented at the Annual Meeting of the American Educational Research Association*. Apr 11-15, 2005. Montreal, Quebec, Canada.
- Mohd Azhar Mat Ali (2006). *Amalan Pentaksiran Sekolah Menengah di Malaysia*. Universiti Malaya. Tesis PhD.
- Muhamad Sahari Nordin (2001). Pengujian Selaku Pemangkin Perubahan Pendidikan : Satu Peluang atau Retorik? *Jurnal Pengurusan Pendidikan Institut Aminuddin Baki*. 11(2), 25-36.
- Nitko, A. J. (2001). *Educational Assessment of Students*. Upper Saddle River, NJ : Merrill.
- Nunnally, J.C. (1978). *Psychometric Theory* (2nd ed.). New York : McGraw-Hill Publishing Company.
- Othman Mohamed (2001). *Penulisan Tesis Dalam Bidang Sains Sosial Terapan*. Penerbit Universiti Putra Malaysia, Serdang.
- Pajares, M.F. (1992). Teachers' Beliefs and Educational Research : Cleaning Up a Messy Construct. *Review of Educational Research*, 62, 307-332.
- Plake, B. S. (1993). Teacher Assessment Literacy : Teachers' Competencies in the Educational Assessment of Students. *Mid-Western Educational Researcher*, 6(1), 21 - 27.
- Plake, B. S. (1993). Teacher Assessment Literacy : Teachers' Competencies in the Educational Assessment of Students. *Mid-Western Educational Researcher*, (1), 21-27.
- Plake, B.S., Impara, J.C., dan Fager, J.J. (1993). Assessment Competencies of Teachers : A National Survey. *Educational Measurement : Issues and Practice*, 12(4), 10-12, 39.
- Popham, W. J. (1990). *Modern Educational Measurement : A Practitioner's Perspective*. Needham Heights, MA :Allyn and Bacon.
- Robiah Sidin (1998). *Pemikiran dalam Pendidikan*. Shah Alam : Fajar Bakti Sdn Bhd.
- Salbiah Ishak (1998). *Amalan Penilaian Formatif Dalam Pelaksanaan Kemajuan Berasaskan Sekolah Oleh Guru Bahasa Melayu dan Matematik*. Tesis Sarjana Pendidikan. Universiti Kebangsaan Malaysia.
- Schaefer, W. D. (1993). Assessment Literacy For Teachers. *Theory Into Practice*, 32(2), 118-126.
- Schaefer, W.D. Dan Lissits, R.W. (1987). Measurement Training for School Personnel : Recommendations and Reality. *Journal of Teacher Education*, 38(3), 57-63.
- Simmons, W. dan Resnick, L. (1993). Assessment as the Catalyst for School Reform. *Educational Leadership*, 50, 11-15.
- Stiggins, R. J. (1991). Relevant Classroom Assessment Training for Teachers. *Educational Measurement : Issues and Practice*, 10(1), 7-12.
- Stiggins, R. J. (1995). Assessment Literacy For The 21st Century. *Phi Delta Kappan*, 77(3), 238-245.
- Stiggins, R. J. (1999). Are you Assessment Literate? *The High School Journal*, 6(5), 20-23.

- Stiggins, R.J. (2001). The Unfulfilled Promise of Classroom Assessment. *Educational Measurement : Issues and Practice*, 20(3), 5-15.
- Title, C.K. (1994). Toward an Educational Psychology of Assessment for Teaching and Learning : Theories, Contexts and Validation Arguments. *Educational Psychologist*, 29, 149-162.
- Webb, N. L. (2002). Assessment Literacy in a Standards-Based Education Setting. *Paper presented at the Annual Meeting of the American Educational Research Association*. April 1-5,2002. New Orleans, Louisiana.